CHAPTER 6

ENVIRONMENTAL DATA COLLECTION AND ANALYSIS

6-1. General Considerations. Most flood control channel projects will not require detailed environmental studies. However, extensive, sensitive, or extremely complex projects may require a more intensive effort. In the process of planning and designing these projects, assessment of potential environmental effects and opportunities requires site-specific data collection efforts. While the details of data collection and analysis are specific to each project, there are basic requirements that are common to all data collection programs. This chapter does not direct data collection efforts but outlines the general steps to be considered when undertaking data collection programs (Figure 6-1). Engineer Manual 1110-2-1201 presents specific information concerning water quality data.

a. Problem Identification. Before objectives for a data collection effort are set, the problem to be addressed must be clearly identified. The general (and sometimes specific) nature of the problem may be ascertained from a variety of sources. These include Environmental Assessments, Environmental Impact Statements, Feasibility Reports, Reconnaissance Reports, consent decrees, statutes, regulations, and interagency agreements.

b. Establishment of Objectives.

(1) Need for objectives. The most essential part of an environmental data collection and analysis effort is the establishment of clear objectives. If this is not done, the net result is often either an inability to solve the problem for which the data were generated or a mass of data that defies rational analysis. Without good objectives, any data collection/analysis effort faces a high probability of failure. The various stages of the project may warrant different details of problem identification and objectives.

(2) Nature of objectives. An approach to setting objectives is presented in Phenicie and Lyons (1973). Objectives must be attainable, oriented in a positive direction with no unproductive branching, and measurable to allow evaluation of progress and results. Wording must be clear, unambiguous, concise, and simple. The types of data needed to measure success or results should be specified.

c. Study Design.

(1) The design describes how objectives will be met and includes decisions on parameter and variable selection, data collection methods, study milestones, resource allocation, and necessary reports. Use of CPM (Critical Path Method) logic networks is often helpful in outlining work to be accomplished and the sequence of events.

(2) Simple before-and-after studies of the project area may be used to document changes but usually are insufficient to establish causal relationships (i.e., observed effects result from specific actions or variables). If the study is to identify cause-and-effect relationships, it is necessary to

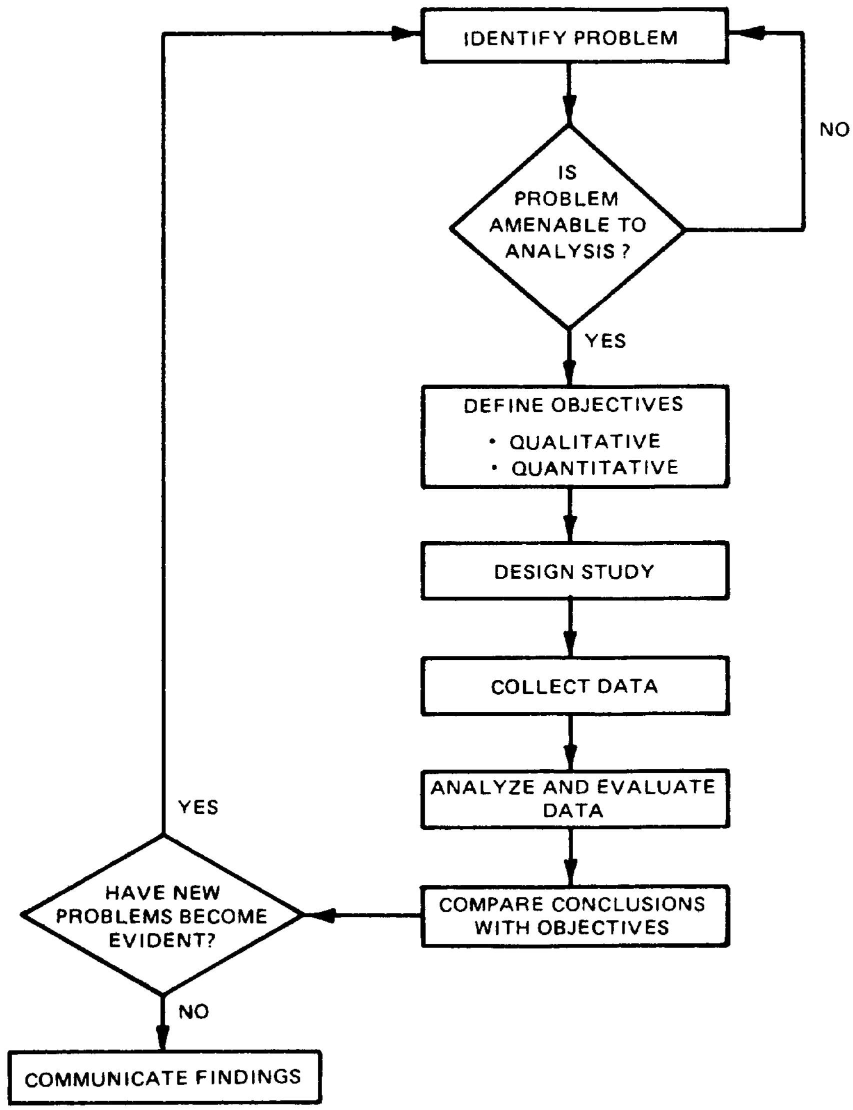

Figure 6-1. Major steps in conducting environmental studies

identify and control for other variables or processes that might influence similar results. Temporal changes in environmental processes can sometimes be accommodated by using control areas within or in the immediate vicinity of the project site. Environmental conditions in control areas should be as similar as possible to those of the project area.

d. Types of Data.

(1) There are two basic kinds of data: qualitative and quantitative. Qualitative data consist of descriptive, nonnumerical information. Quantitative data are numerical and usually reference temporal or spatial information. Qualitative approaches are especially useful if only descriptive data are required, if the study is preliminary in nature, if a short suspense has been set, or if quantitative techniques do not apply. Quantitative data are

preferable because they can be expressed as a testable hypothesis. It is often useful to express the hypothesis as a question, for example, "Will (has) the project increase(d) (decrease(d)) some variable" (e.g., density)? The objective of data collection and analysis then essentially becomes the verification or rejection of a hypothesis.

(2) For scientifically and legally defensible conclusions, baseline monitoring and reference data should be quantitative and reproducible and the experimental design such that hypotheses concerning change can be statistically evaluated. Quantitative data sufficient for application of statistical tests are often quite expensive, a fact that underlines the importance of careful selection of parameters for measurement.

e. Documentation. Documentation of study findings is critical to the future use of the environmental data collected. Reporting requirements should be incorporated into the study design, taking into consideration the report format to be used. A common format used in reports of results consists of the following parts: Introduction, which contains background information, the problem, and how specific objectives will lead to resolving the problem; Materials and Methods, which includes a description of the study area and detailed field and/or laboratory procedures, sampling techniques, and methods for analyzing the data; Results, which gives measurements of variables and results of hypothesis testing; Discussion, which presents and explains the results; Conclusions and Summary; and Literature Cited.

6-2. Data Collection. This section provides guidance for planning a sampling program that will meet stated objectives of the study design. The most critical aspect of data collection is the selection of appropriate parameters to sample and measure.

a. Primary Considerations. The quality of information obtained through the sampling process is dependent upon collecting a representative sample and using appropriate sample collection and data management techniques. Time, costs, and equipment constraints may limit the amount of information that can be gathered. Under such conditions, careful tailoring of the data collection program is required. In this document, the term sample refers to a set of observations or measurements taken to describe or characterize environmental conditions. Individual observations or measurements are called sample elements, and the number of sample elements constitutes the sample size.

b. Representative Sampling. The purpose of sampling is to define biological, physical, or chemical characteristics of the project area environment. This requires that samples be taken from locations that are typical of ambient conditions found at the project site. Failure to obtain samples that are truly representative of a given location will result in inaccurate data and misinterpretations. Samples can be random, haphazard, or stratified and will be specified in the sampling design. Elliot (1977) and Green (1979) provide information on these aspects.

c. Sampling Site Selection and Location. The following general factors should be considered in selecting a sampling site:

(1) Objectives of the study.

(2) Accessibility of the site to personnel and equipment.

(3) Representativeness of the site.

(4) Available personnel and facilities.

(5) Other physical characteristics.

Statistical texts and field manuals in geology, hydrology, biology, and other environmental disciplines should be consulted for information about specific factors to consider in sample site selection.

d. Sample Size. Guidance in this section is limited to general concepts. The larger the sample size, the better environmental conditions will be defined. The mean of a series of replicated measurements is generally less variable than a series of individual measurements. Statistical analysis generally requires at least two characteristics, usually mean and standard deviation, to describe a sample. The sample size necessary to describe a distribution is proportional to the heterogeneity of the variable to be measured. Refer to Snedecor and Cochran (1967), Elliot (1977), Green (1979), or other statistical texts for numerical and graphical techniques to determine sample size.

(1) Consideration of the above factors suggests that replicate samples should be collected when money and time permit. A minimum of three replicates is required to calculate standard deviations. Aside from replication, the sample size needed depends on temporal and spatial variability of the phenomenon and the desired degree of precision.

(2) An additional factor that will limit the sample size is financial resources. In this case, the sample size that can be analyzed is determined by the ratio of available dollars and cost per observation:

$$\text{Sample size} = \frac{\text{Dollars available}}{\text{Cost per observation}}$$

This approach will provide one method of estimating the sample size that can be collected and analyzed. However, should the calculated sample size be insufficient to establish an adequate sampling program (i.e., sample size insufficient to allow replicate measurement at all locations), one of the following trade-offs will have to be accepted:

(a) Reduce the replicate sampling at each station.

(b) Maintain replicate sampling but reduce the number of sampling locations.

(c) Increase the financial resources available for sampling.

The distinction between options (a) and (b) above should be based on project-specific goals. If option (a) is selected (more stations, fewer replicates), the results will provide a better indication of distribution patterns in the project area (synoptic survey), but it will be difficult to compare individual stations. On the other hand, if option (b) is selected (fewer stations, more replicates), the results will provide a better indication of variability at one location and comparison between sampling stations. However, the project area will be less described.

(3) Consideration should be given to collecting a larger sample than that determined by the above process in the event they are needed for additional analysis or backup. If more data are needed, it is easier to analyze additional sample elements already on hand than to remobilize a field crew. Also, the additional variable of different sampling times is avoided with this approach.

e. Sampling Collection, Handling, and Analysis. Refer to EM 1110-2-1201 and references contained therein for information concerning collection, handling, and analysis of water, sediment, and biological samples.

6-3. Data Analysis, Interpretation, and Presentation of Results.

a. Data Analysis Plan. A plan for data analysis should be formulated at the experimental design step since the type of analysis selected will guide the sample size and frequency of measurements that must be taken. Techniques available for data analysis include descriptive analysis, maps and graphical analysis, and statistical analysis.

(1) Descriptive analysis. Presentation of the results of some analyses often consists of descriptions based on visual observations, inductive reasoning, and perhaps a few measurements: for example, "The habitat structures placed in the flood channel have provided flow depth and velocity variation. Prior to placement of the structures, water depths and flow velocities were relatively uniform across and along the channel. After the structures had been in place for 6 months, scour holes several feet deep had developed at the riverward end of each structure." The value of descriptive analysis can be substantial if it can be established that other factors that could affect results were controlled, constant, or not applicable.

(2) Maps and graphical analysis. Patterns inherent in data can often be revealed by mapping or graphing the data. Maps are used to show two- and three-dimensional spatial patterns, whereas graphical approaches are most useful for showing temporal relationships or variations within a single dimension, such as distance or depth.

(a) Maps. Phenomena to be mapped may be distributed in a continuous or discrete manner. Discrete distributions are composed of individual elements that are countable or measurable (such as people, fish, or trees), whereas continuous distributions have no recognizable individuals (e.g., air temperature or rainfall). Patterns are often enhanced by grouping all values into five or six classes and mapping each class with a separate tone or color.

(b) Graphs. Graphic techniques specialized for certain disciplines or types of data are too numerous to describe. As with maps, however, graphic techniques vary with the type of data. Discrete data are often graphed as frequency histograms, with frequencies on the vertical axis and classes or categories on the horizontal axis. More complex maps and graphs, such as three-dimensional contour plots, trend surfaces, and perspective plots, are also useful but more difficult to comprehend. Various mapping and graphic options are available as part of most data management systems. Continuous data are usually plotted as curves, with the spatial or temporal dimension on the x-axis and the values of the variables on the y-axis.

(c) Common errors. When using maps and graphic techniques, one must be careful not to draw conclusions that implicitly depend on interpolation between data points (Figure 6-2) or extrapolation beyond the range of the data (Figure 6-3), unless such interpolation or extrapolation can be justified. A choice of scales or coordinate axes that unduly exaggerate or minimize point scatter or differences should be avoided.

(3) Statistical analysis.

(a) General. Statistical analysis can be used to summarize or describe data. Statistics can also be used as a formal decision-making tool to decide whether measured temporal or spatial differences between samples are real or whether they may be the result of sampling variability. Commercially available data management software systems have options for computing and displaying several types of statistics. Large amounts of data can be summarized by calculating statistics such as measures of central tendency (mean, median, and mode) and dispersion (standard deviation and range). Statistics can be used to compare sets of data to determine if differences exist among them and, if so, whether the differences are meaningful.

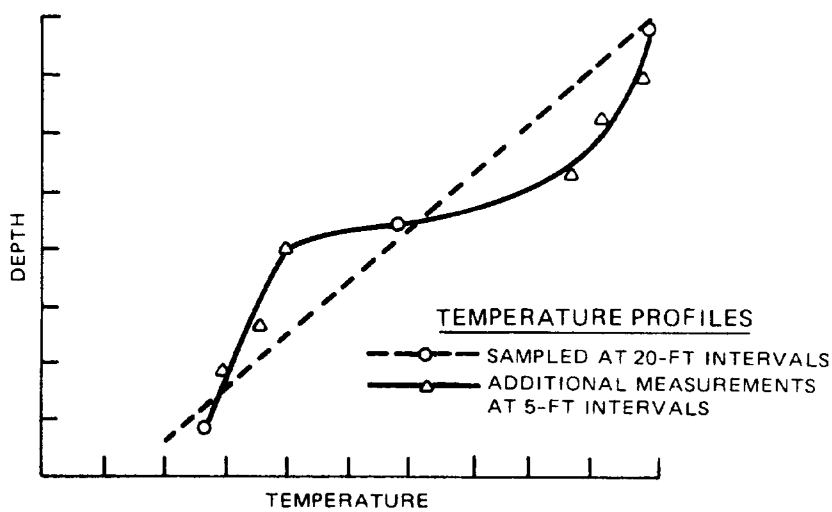

Figure 6-2. Error caused by improper interpolation. Depth-temperature relationship appears linear when sampled at 20-foot intervals, but non-linear when sampled at 5-foot intervals

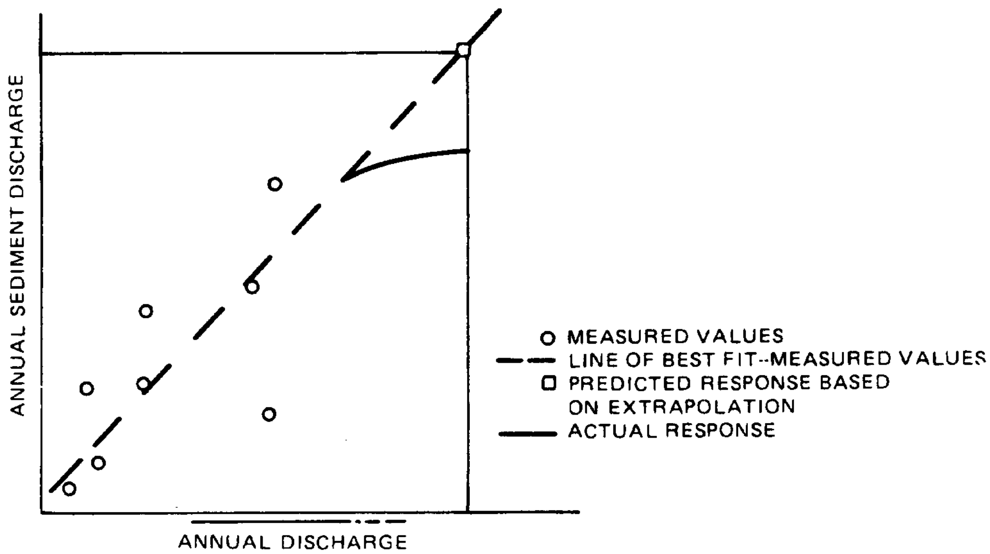

Figure 6-3. Error caused by improper extrapolation

(b) Hypothesis testing. Formulas are available for determining if observed differences between sample data sets are real, or if they may have occurred by chance due to the size or selection of samples used in calculating the statistics. These techniques are called significance tests, and theories and formulas for their use are given in basic texts on statistics and experimental design. Users should be cautioned, however, that observed differences may be statistically significant and yet not be very meaningful. Special techniques have been developed for analysis of biological data, particularly benthic data. Sokal and Rohlf (1969) provide a review of several of these techniques.

b. Data Interpretation.

(1) Editing. Data base checking and editing should precede analysis. Extreme errors may be detected by computer programs that check for boundary conditions and ensure that data values are within reasonable limits. Quality work requires human judgment. Simple computer-generated plots of the raw data should be examined for unreasonable values, extreme values, trends, and outliers.

(2) Analysis. The next step in data interpretation is to ensure that the assumptions upon which the data analysis plan is based are still valid. New information or failure to collect all the data required by the original analysis plan may necessitate modification. The final conclusions should not be limited to acceptance or rejection of hypotheses, but should extend to clear, verbal expression of the implications of the observed results. Decision-makers who are not technical specialists may fail to grasp these implications unless they are clearly communicated.

c. Presentation of Results. Results should be presented in a format appropriate for the majority of the intended audience. Presentation of large volumes of numerical data is generally undesirable; however, provision should be made for long-term data storage and retrieval (computer disks, microfiche, etc.). Graphic displays can effectively serve as examples of major findings or conclusions.

6-4. Data Base Management.

a. General. The success of any study effort, especially one involving multiple investigators and disciplines, will be heavily influenced by the quality of data management, storage, and efficiency of information retrieval and by the compatibility between data units and the formats and programs for data reduction and analysis. A carefully designed plan for handling information will guarantee that once field and laboratory work are completed, information will be readily available for examination and analysis, in a form useful to management.

b. Data Management Plan. A data management plan detailing procedures for handling data storage and retrieval should be formulated at the outset of an environmental study. The simplest type of data base contains only data developed for a single study. Efforts should be made to ensure standardization of measurement and reporting procedures so that there will be internal compatibility among the environmental data files within the Corps field office. Once the data base is developed, the data base manager should be conservative in decisions about changes in procedures or data units and should permit such changes only where useful information benefits can clearly be identified.

c. Data Base Incompatibility. Frequently, various studies associated with one project will be conducted by several different agencies or contractors. The same scope of work might be performed by different contractors at different times. Besides reinforcing the need for standardization, the probability of a multiple-contractor operation brings up logistical questions about information storage, retrieval, and analysis. Federal agencies, academic institutions, and consulting companies who ordinarily conduct Corps contracts will usually have their own computer support. This situation could lead to the formation of incompatible data files. Data base incompatibilities will create problems for those who have responsibility for synthesizing the products of multiple investigators and will hamper comparisons over time. A solution is to permit each contractor to use the computer hardware and software of his choice, but to require the contractor to transmit information to the Corps field office in a machine-readable form compatible with Corps format and units.